The Moving Crowd:

Collecting and Processing of Crowd Behaviour Data

Rik Bellens
PhD, Department of Telecommunication and Information Processing, Ghent University
St. Pietersnieuwstraat 41, B-9000 Ghent, Belgium
T. +32 9 264 34 10, E. Rik.Bellens@UGent.be
Sven Vlassenroot
PhD, Department of Telecommunication and Information Processing, Ghent University
St. Pietersnieuwstraat 41, B-9000 Ghent, Belgium
T. +32 9 264 34 10, E. Sven.Vlassenroot@UGent.be
David Verstraeten
Dr., Reservoir Lab, Ghent University
St. Pietersnieuwstraat 41, B-9000 Ghent, Belgium
T. +32 9 264 34 04 , E. David.Verstraeten@UGent.be
Sidharta Guatama
Prof. dr., Department of Telecommunication and Information Processing, Ghent University
St. Pietersnieuwstraat 41, B-9000 Ghent, Belgium
T. +32 9 264 34 10, E. Sidharta.Guatama@UGent.be

ABSTRACT

The MOVE project focuses on the collection and analyses of crowd behavior data. The two main goals of the project are first, the collection of data through mobile phones. The second goal is to develop new technologies to process and mine the collected data for crowd behaviour analysis. The technology will allow to make advanced interpretations of historic and dynamic mobile crowd data coming from GSM/GPS and from different classes of users (vehicle, pedestrian, indoor/outdoor). Fusion will be made between data coming from different sources (smartphone, navigation device) and external map data. The interpretation will allow the mining of advanced features/geometry from the crowd data as well as interprete the dynamic behaviour of the population.

INTRODUCTION

The sector of wireless and mobile telecommunication is expected to play a pivotal role. Recent technological developments have produced a range of digital tracking technologies that offer a view on the movement of users, which has given rise to **location based services (LBS)**. Tracking technology can be integrated with current mobile phones and PDA's. The simple and standard solution is GPS-based devices. Less known is the fact that mobile phones also offer the possibility to track people constantly. Operating on a phone network requires the network operator to be able to constantly detect the subscriber's proximity to a specific antenna, even when no calls are made. In general, the accuracy of tracked mobile devices is lower than GPS-devices, ranging from 50 to 100m. Projects like MIT's Senseable City investigate behaviour patterns through cell phone activity. The analysed activity is still limited to presence detection within a single cell tower range (typical resolution 100m) and does not take into account dynamic spatial movement patterns. However, activity during large events over the city of Amsterdam are clearly observed

The MOVE project pushes this technology into its next generation: UGent has specialized processing algorithms which enhances the bundle of collected position tracks from GPS or GSM through map matching and data mining. Enhancement is achieved of a full order of magnitude (e.g. 50m to 5m for GPS in urban canyons). More importantly, topology and movement behaviour is retained which allows dynamic models for crowd behaviour to be applied on the data. The extracted information on crowd behaviour can service markets like market research, intelligent transportation systems (ITS) and security.

THE MOVE RESEARCH SETUP

The MOVE project has two main goals. The first is the collection of data through mobile phones. The second goal is to develop new technologies to process and mine the collected data for crowd behaviour analysis.

DATA COLLECTION

In order to collect data for crowd behaviour analysis, we have developed a mobile software application running on Android phones and Java Micro Edition. This application collects different kinds of valuable information and sends it to a central server. If the cell phone contains a gps chip, accurate locations of the phone can be collected. However, because gps is very demanding on the battery and does not work inside buildings, the gps is only used by the software when most appropriate. At other times, other information is used to derive the location of the phone as accurately as possible, such as the current and neighbouring cell towers, the wifi stations in sight and their signal strengths. On top of that, we also collect

measures of the accelerometer of the phones to help us distinguish between pedestrians, cyclists and cars. Currently, a group of students of the University of Ghent have installed this software and are constantly collecting data in and around Ghent, Belgium. Figure 1 shows the data that has been collected during the first two weeks of the trial.

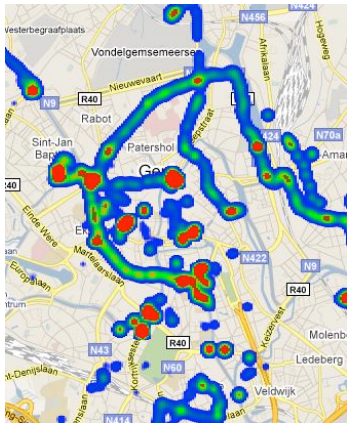

Figure 1. Location data collected during the first two weeks of the trial.

DATA PROCESSING TECHNOLOGY

The location of individual phones in itself is not very useful. It only becomes interesting when the locations of many users together can be analyzed. However, going from these individual locations to crowd behaviour information, is still a large gap to bridge. Simple statistical analysis of the data can already be interesting, however, the truly interesting information is hidden and sophisticated technologies are needed to analyze the data. Since, the positional data derived from wifi and cell information is very inaccurate, the algorithms have to take into account this inaccuracy.

The proposed workflow to process the data consists of three stages:

- In a first stage, the accuracy is enhanced and the geometry is extracted through a multidimensional rastering and fuzzy voting mechanism. This allows us to extract even the fine details of movement flows (see figures 2 and 3).
- The second stage uses error tolerant graph matching to match the data with existing vector data such as road maps.
- Thirdly, different local features are calculated on the data and data mining techniques, such as neural networks, are used to derive high quality crowd behaviour data.

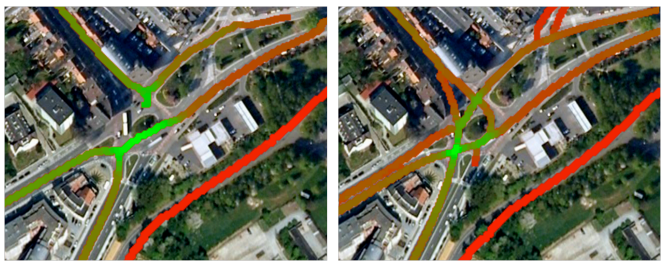

Figure 2. Examples of deriving geometry from gps location data.

Two different methods are compared. On the left, the result of a two-dimensional rasterizing method is shown. The different flows cannot be distinguished here. On the right, a more dimensional rasterizing method is used. Here, all the different routes are extracted.

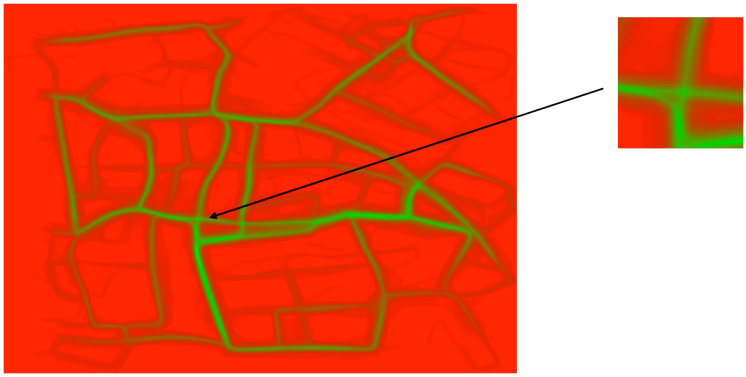

Figure 3. Example of the enhancement of accuracy through multidimensional rastering with fuzzy voting.

The detail of the cross-road shows that even fine details become visible.

MODELING MOBILE CROWD BEHAVIOUR

Individual position tracks as such hold limited information. Reports have to be generated that summarize critical aspects of crowd behaviour. UGent has specific know-how in the field of socio-economic geography and mobility, where the impact of behaviour patterns on transport, mobility and urban development is studied. The patterns are agent based, meaning that the complex behaviour of groups of people are modelled as the interaction of autonomous agents. This leads to a much more realistic and understandable model in terms of classes of user profiles.

CONCLUSION

The technology will allow to make advanced interpretations of historic and dynamic mobile crowd data coming from GSM/GPS and from different classes of users (vehicle, pedestrian, indoor/outdoor). Fusion will be made between data coming from different sources (smartphone, navigation device) and external map data. The interpretation will allow the mining of advanced features/geometry from the crowd data as well as interprete the dynamic behaviour of the population.

The technology has a broad market potential which will be explored within the project. The project will work towards **creating value in the field of ITS, geomarketing and security**. In addition, research will benefit from the large-scale data that will be collected during the project and additional funding which will be raised.